算法治理对数字平台生态系统的重构效应

——基于微博热搜榜的准自然实验

摘　要：算法推荐系统作为数字平台的核心信息配置机制，在优化用户体验的同时也引发了信息茧房、注意力垄断等公共治理难题。如何在释放算法价值的同时有效规制其负外部性，成为全球数字治理的核心议题。本文利用 2024 年 11 月中国清朗行动算法治理政策这一准自然实验，采用断点回归时间序列与双重差分相结合的识别策略，基于微博热搜榜高频数据系统评估算法治理对平台内容生态的重构效应。研究发现，算法治理有效推动了注意力分布的去中心化，瓦解了头部话题的垄断格局；分级分类治理导致社会类内容质量显著提升而娱乐类保持稳定，形成差异化的质量改善路径。然而，在注意力总量守恒的零和约束下，监管对高风险领域的强化治理引发了注意力向低风险领域的系统性转移，呈现典型的“水床效应”。在密度层面，社会类呈“少而精”调整，娱乐类呈“多且长”扩张。本研究揭示了算法治理的复杂效应：多样性提升与质量改善的积极成效，与注意力再分配的非预期后果并存，形成“监管初衷”与“生态演化”之间的张力。研究为理解算法治理的系统性后果提供了新的实证证据，对完善分级分类治理体系、探索“激励优质内容 + 动态监测调整 + 跨平台协同治理”的系统性框架具有重要政策启示。

关键词：算法治理；平台生态系统；注意力经济；断点回归时间序列；水床效应
JEL 分类号：D83；L51；L86

Abstract: While algorithmic recommendation systems optimize user experience, they also trigger concerns about information cocoons and attention monopolization. How to effectively govern algorithms has become a core issue in global digital governance. This study leverages the quasi-natural experiment of China's Qinglang Action algorithm governance policy implemented on November 12, 2024, employing interrupted time series (ITS) combined with difference-in-differences (DID) methods. Based on high-frequency data covering 608 days from February 2024 to October 2025 with over 253,227 observations from Weibo's trending list, we systematically evaluate the restructuring effects of algorithm governance on platform ecosystems.

We find: First, algorithm governance significantly promotes attention distribution decentralization, with Shannon entropy immediately increasing by 18.7% (level coefficient +0.146, $p < 0.001$) and HHI decreasing by 11.8% (level coefficient -0.0292, $p < 0.001$). The platform breaks the head monopoly pattern by limiting single-topic repeated listings (down 23.4%) and

continuous listing duration (shortened by 31.7%). Second, risk-based tiered governance leads to content quality differentiation. Social topics' information gap index decreases by 0.561 at level ($p = 0.036$), with official media showing a net decrease of 0.0205 ($p < 0.001$) relative to non-official media in information gap, while entertainment topics show no significant quality changes. Third, under zero-sum attention constraints, regulation triggers systematic attention reallocation from social to entertainment content. Social topics' share immediately declines by 3.16 percentage points ($p < 0.001$) with continuing downward trends, while entertainment rises by 3.84 percentage points (celebrity +3.61pp, gaming +1.12pp). Weekly total heat discontinuity tests verify total conservation ($p = 0.63$), exhibiting a pronounced "waterbed effect." Fourth, attention density significantly differentiates across categories, with social topics showing "fewer but refined" (average listing duration per topic significantly increases) and entertainment core categories showing "more and prolonged" (total listing duration, time share, and average listing duration all significantly increase).

This study proposes theoretical mechanisms of "meta-organization of meta-organizations" and "waterbed effect," extending platform governance theory boundaries. The research reveals complex effects of algorithm governance: while regulation successfully achieves diversity improvement and social content quality enhancement, it triggers unintended attention transfer under zero-sum attention constraints. The large-scale reallocation from social to entertainment content may weaken public issue exposure, creating tension between "regulatory intent" and "ecosystem evolution." The study provides new evidence for understanding systemic consequences and unintended effects of algorithm governance, offering important policy implications for improving tiered governance systems and constructing a systematic framework of "quality incentives + dynamic adjustment + cross-platform coordination + user empowerment."

Keywords: Algorithm governance; Platform ecosystems; Attention economy; Interrupted time series; Waterbed effect
JEL Classification: D83; L51; L86

一、引言

数字平台已成为信息传播、社会互动和经济活动的核心基础设施。作为平台生态系统的“神经中枢”(Jacobides et al., 2018)，算法推荐系统在优化用户体验、提升商业价值的同时，也在深刻塑造着公共信息环境。当算法推荐逐渐成为亿万用户获取信息的“把关人”(Gorwa et al., 2020)，其带来的信息茧房、注意力垄断、内容同质化等问题日益凸显，如何有效治理算法推荐系统成为全球数字治理的核心议题。

中国在算法治理领域走在世界前列。自 2021 年《互联网信息服务算法推荐管理规定》发布以来，监管部门持续推进算法治理实践，形成了以“清朗行动”为代表的分级分类治理体系。2024 年 11 月 12 日，国家网信办启动新一轮清朗行动，聚焦算法推荐治理，发布《算法推荐服务专项治理清单指引》，明确 27 项核验标准，要求各大平台“不得利用算法操纵热点话题”、“保障信息内容多样性”、“降低低俗内容权重”，并强化分级分类管理，对社会类等高风险内容实施更严格的审核机制。这一政策冲击为学术界提供了难得的准自然实验场景：当外部监管力量介入平台算法系统，生态系统将如何响应？监管目标能否实现？又会产生何种非预期后果？

现有文献在理论和实证两个维度存在明显不足。理论层面，尽管平台治理研究已积累丰富成果 (Huber et al., 2013; Parker and Van Alstyne, 2018; Tiwana et al., 2010)，但既有研究主要关注平台自身的治理实践，较少探讨外部监管如何触发生态系统的内生调整。算法治理文献 (Gritsenko and Wood, 2022; Ulbricht and Yeung, 2022) 多停留在规范性讨论，缺乏基于经济学框架的微观机制分析。特别是，在注意力作为稀缺资源的零和竞争环境下，对某一内容类型的监管强化如何引发注意力在不同类型间的系统性再分配，现有研究尚未提供清晰的理论解释。实证层面，由于算法系统的“黑箱”特性和高频政策调整，学界难以获得清晰的政策断点和高质量数据，导致算法治理效果的实证研究极为稀缺 (Jhaver et al., 2019)。

本文聚焦 2024 年 11 月清朗行动算法推荐治理这一准自然实验，以微博热搜榜为研究对象，采用断点回归时间序列与双重差分相结合的识别策略，基于覆盖政策前后共计 608 天、超过 25 万条观测值的高频数据，系统评估算法治理对平台内容生态的多维重构效应，揭示监管在改善内容生态的同时引发注意力结构性再分配的复杂图景，为理解算法治理的系统性后果和非预期效应提供实证证据。

本文的主要贡献体现在以下三个方面。

理论层面，本文拓展了平台治理理论的分析边界。不同于既有研究关注平台内部的治理机制设计 (Boudreau and Hagiu, 2009; Huber et al., 2013)，本文引入“元组织的元组织”概念 (Jacobides et al., 2018)，揭示政府如何通过算法监管成为平台生态系统的上层治理者，通过改变平台的约束集合（多样性下限约束、内容质量约束、透明度约束）触发生态系统的内生调整。本文提出“水床效应”机制，系统化地解释了在注意力零和约束下，对某一内容类型的监管强化如何导致注意力向其他类型溢出，形成生态系统的级联调整。具体而言，当监管大幅提升社会类内容的审核成本和质量门槛时，平台为维持商业价值最大化，会将受挤压的注意力配额再分配至审核成本较低的娱乐类内容，形成“社会类紧缩—娱乐类扩张”的跷跷板效应。这一机制丰富了算法治理的理论内涵，为理

解监管干预的系统性后果和非预期效应提供了新视角。

实证层面，本文是国内首次利用准自然实验评估算法治理效果的研究。研究克服了算法系统“黑箱”和数据获取难题，构建了覆盖 608 天、包含多维度指标（多样性、质量、注意力配置、注意力密度）的微观面板数据。通过严谨的 ITS 和 DID 设计以及多重稳健性检验（安慰剂检验、平行趋势、异质稳健标准误、替代指标、断点敏感性、排除时间趋势），本文提供了算法治理因果效应的可信估计，为政策评估文献增添了中国证据。研究发现，算法治理在实现多样性提升（Shannon 熵增加 18.7%，HHI 下降 11.8%）和社会类内容质量改善（官媒相对非官媒信息差净下降 0.0205）的同时，也带来了非预期的注意力转移——社会类份额即时下降 3.16 个百分点且趋势项持续走低，娱乐类份额上升 3.84 个百分点（明星类 +3.61pp，游戏类 +1.12pp），周度总热度守恒验证（$p = 0.63$）印证了注意力零和约束的存在，揭示了监管政策的复杂动态效应。

政策层面，研究结果对完善算法治理体系具有重要启示。研究发现，尽管分级分类治理在短期内有效改善了信息环境（多样性提升、社会类质量改善），但注意力零和约束使得监管目标面临“按下葫芦浮起瓢”的困境——社会类份额的下降可能削弱公共议题曝光度，娱乐类份额的上升可能重塑用户信息消费结构，形成“监管初衷”与“生态演化”之间的张力。这提示监管部门需要采取更加系统化的治理策略：第一，从“压制劣质”转向“激励优质”，建立社会类优质内容的识别和激励机制，通过正向引导而非单纯限制实现内容质量提升；第二，从“静态规则”转向“动态调整”，建立基于平台生态反馈的动态监管机制，及时识别和纠正非预期后果；第三，从“单平台治理”转向“跨平台协同”，防止注意力在不同平台间的溢出和套利行为；第四，从“平台约束”转向“用户赋权”，通过算法透明化和用户控制权增强，让用户成为信息多样性的主动塑造者。

本文的结构安排如下：第二部分构建理论框架，推导研究假说；第三部分介绍研究设计和数据来源；第四部分报告实证结果及稳健性检验；第五部分讨论研究发现的理论含义和政策启示。

二、制度背景、文献回顾与研究假设

（一）理论基础：注意力作为稀缺资源的平台配置逻辑

1. 注意力的零和性质与竞争机制

在数字平台生态系统中，注意力构成了比信息本身更为稀缺的核心资源 (Oestreicher-Singer and Sundararajan, 2012)。当内容供给呈指数级增长时，用户的认知带宽却受制于生理和时间约束而保持刚性。根据新浪微博 2023 年财报，用户日均使用时长约 52 分钟，其中热搜浏览时长约 15–20 分钟，这一有限的时间窗口构成了注意力资源的刚性约束。

注意力资源具有三个决定性特征。第一，稀缺性：用户日均使用时长存在上限，短期内缺乏弹性 (Hosanagar et al., 2014)。第二，竞争性：注意力分配给某一话题必然减少其他话题的份额 (Berman and Katona, 2020)。在任意时刻，热搜榜只显示 50 个固定位置，某一话题进入榜单必然挤出其他话题，形成典型的零和博弈格局。第三，不可再生性：注意力一旦消耗无法恢复，用户在某条娱乐话题上停留的时间永久性地从社会类话题的潜在配额中流失。

这种竞争关系在社交媒体平台上表现为注意力拥挤效应：当信息供给增加时，单位

信息获得的注意力边际递减 (Gelper et al., 2018)。设平台上有 N 条话题竞争用户注意力，用户总注意力预算为 $\bar{A}$，则第 i 条话题获得的注意力 A_i 满足注意力守恒约束：

$$\sum_{i=1}^{N} A_i = \bar{A} \tag{1}$$

这一效应在热搜榜体系中表现为榜内与榜外话题的阅读量呈数量级差异，榜单位置的瞬时稀缺性使得话题间竞争异常激烈。更为关键的是，热搜榜不仅是注意力的消费场所，更是注意力的生产场所。平台通过算法排序和榜单展示，将分散的用户浏览行为转化为可度量、可累积的社会关注度 (Lorenz-Spreen et al., 2019)。上榜本身构成一种注意力增值过程，话题进入榜单后获得的注意力呈指数级增长，形成正反馈循环。因此，热搜榜单的任何调整，都是对稀缺注意力资源分配权力的重新定义。

2. 平台的约束优化模型

基于注意力零和博弈的底层逻辑，平台内容分发系统可被抽象为一个多目标约束优化问题。传统上，社交媒体平台以最大化用户参与度为首要目标 (Hagiu and Wright, 2015; Kleinberg and Raghavan, 2021)。设平台上第 i 类内容获得的注意力为 A_i，其产生的用户参与度为 $R_i(A_i)$，平台的目标函数可表示为：

$$\max_{\{A_i\}} \quad \sum_{i} R_i(A_i) \tag{2}$$

其中，$R_i(A_i)$ 通常为凹函数，体现了用户参与度的边际递减特征 (He et al., 2022)。在监管介入前，平台仅面临基本的资源约束 $\sum_i A_i = \bar{A}$，这一优化问题的解往往导致注意力高度集中于少数头部内容，因为头部内容通常具有更高的边际参与度 $R_i'(A_i)$。

算法治理行动根本性地改变了平台的约束集合。Huber et al. (2013) 指出，外部监管通过引入新的治理约束，改变了平台生态系统价值创造与治理成本之间的权衡。具体而言，《算法专项治理清单指引》引入了三类新约束：

第一，多样性下限约束。设 $D(\{A_i\})$ 为衡量注意力分布多样性的指标（如 Shannon 熵），监管要求：

$$D(\{A_i\}) \geq D_{\min}^{\text{post}} > D_{\min}^{\text{pre}} \tag{3}$$

这一约束直接限制了头部内容的注意力集中度，迫使平台为长尾内容预留基础曝光配额。Song et al. (2018) 的研究表明，平台治理机制的动态调整往往是为了满足外部利益相关者施加的约束，而非自发的优化行为。

第二，内容质量约束。设 $C_i(Q_i, R_i)$ 为维持第 i 类内容质量水平 Q_i 的治理成本，该成本函数依赖于内容的风险等级 R_i。监管要求：

$$Q_i \geq Q_i^{\min}(R_i), \quad \frac{\partial Q_i^{\min}}{\partial R_i} > 0 \tag{4}$$

对于风险较高的社会类内容，$Q_i^{\min}$ 显著提升，相应的审核成本 C_i 呈超线性增长 (Shin, 2021)。

第三，透明度与可追溯性约束。监管要求平台留存榜单日志、公示排序机制，实质上增加了平台的合规运营成本。这些新增约束将平台的优化问题转化为：

$$\max_{\{A_i,Q_i\}} \quad \sum_i R_i(A_i) - \sum_i C_i(Q_i, R_i) \tag{5}$$

$$\text{s.t.} \quad \sum_i A_i = \bar{A}, \quad D(\{A_i\}) \geq D_{\min}^{\text{post}}, \quad Q_i \geq Q_i^{\min}(R_i) \tag{6}$$

这一约束优化框架揭示了算法治理的核心机制：通过缩小平台的可行域，监管改变了最优解的结构，进而触发注意力资源的系统性重新配置。

（二）研究假说

1. 假说 H1：注意力去中心化效应

1. 数学推导

命题 1.1（可行域收缩与约束硬化）：算法治理行动改变了平台决策的约束条件集合。设平台内容分发策略空间为 $\mathcal{P}$，监管前平台面临基于用户点击率的软约束，监管后《治理清单》第 7 条将多样性要求转化为硬约束。定义 $H(p)$ 为衡量话题分布多样性的 Shannon 熵，则监管导致的多样性下限 δ 发生位移：

$$\text{s.t. } H(p^{\text{post}}) \geq \delta_{\text{post}} > \delta_{\text{pre}} \tag{7}$$

该约束强化导致平台可行域收缩，剔除了虽然商业价值最大化但熵值低于 δ_{post} 的极点解。

命题 1.2（分布的优劣关系与头部去中心化）：定义 N 个话题的注意力分布向量为 $p = (p_1, p_2, \ldots, p_N)$，并按降序排列 $p_1 \geq p_2 \geq \cdots \geq p_N$。监管措施通过内容去重与打散干预强制削减前 k 个头部话题的流量配额，并将其再分配至长尾话题。数学上，这等价于监管后的分布 p^{post} 被监管前的分布 p^{pre} 所优势，记作 $p^{\text{pre}} \succ p^{\text{post}}$。该关系满足洛伦兹优势条件：

$$\begin{cases} \displaystyle\sum_{i=1}^{m} p_i^{\text{pre}} \geq \sum_{i=1}^{m} p_i^{\text{post}}, & \forall m \in \{1, \ldots, N-1\} \\ \displaystyle\sum_{i=1}^{N} p_i^{\text{pre}} = \sum_{i=1}^{N} p_i^{\text{post}} = 1 \end{cases} \tag{8}$$

此不等式组表明，在任意截断点 m，监管前的累积注意力集中度均不低于监管后。

定理 1（*Shannon* 熵的舒尔凹性）：当注意力分布从高度集中转向相对均匀时，Shannon 熵单调递增。Shannon 熵定义为：

$$H(p) = -\sum_{i=1}^{N} p_i \ln p_i = \sum_{i=1}^{N} \phi(p_i) \tag{9}$$

其中 $\phi(x) = -x \ln x$。计算 $\phi(x)$ 的二阶导数：$\phi''(x) = -1/x < 0$，在定义域 $x \in (0, 1]$ 上恒成立，故 $\phi(x)$ 为严格凹函数。根据优序理论，若 $p^{\text{pre}} \succ p^{\text{post}}$，则必然蕴含 $H(p^{\text{post}}) > H(p^{\text{pre}})$。

2. 理论机制解释

*机制 1.1：目标函数重构——从单目标优化到约束优化。*监管介入前，平台的决策模型可抽象为无约束的流量最大化问题。由于头部话题通常具有更高的商业变现效率，平台倾向于产生极度偏斜的分布。监管实质上引入了拉格朗日乘数 λ，新目标函数为：

$$\mathcal{L}(p, \lambda) = \sum_{i=1}^{N} v_i p_i + \lambda \left(H(p) - \delta_{\text{post}} \right) \tag{10}$$

其中影子价格 $\lambda > 0$ 代表平台为满足合规要求所必须放弃的边际商业收益。Oestreicher-Singer and Sundararajan (2012) 的研究表明，推荐系统能够显著改变产品需求的分布结构，当平台调整算法以提升长尾产品的可见性时，需求分布会从高度集中向更均匀的方向移动。这一机制表明，多样性提升并非市场自发演化的结果，而是通过制度设计强制改变了帕累托最优解的对偶性质。

*机制 1.2：资源编排逻辑——从马太效应到累进调节。*推荐算法的正反馈机制导致了注意力分配的马太效应。《治理清单》中的去重与打散构成了双向调节机制，其作用机理类似于税收制度中的累进税：对高流量话题征收“关注度税”，并转移支付给长尾话题。这种非线性的干预手段切断了“高曝光–高点击–更高曝光”的循环链路，强制系统熵值向最大熵状态漂移。Berman and Katona (2020) 通过理论模型发现，当推荐系统提升内容多样性时，虽然短期内可能降低用户参与度，但长期而言能够提升用户满足度和平台健康度。

*机制 1.3：长尾激活——从被动沉默到主动参与。*在监管前的自然演化状态下，大量长尾话题由于缺乏初始曝光而陷入沉默状态，即使其潜在受众规模不小，也难以突破注意力门槛。算法治理通过为长尾内容预留基础曝光配额，降低了话题进入榜单的边际成本。这种保底机制激活了原本沉默的长尾话题，使其获得接触目标受众的机会，进而可能形成自我维持的关注度增长。

3. 假说陈述

基于上述数学推导与理论机制，提出如下假说：

H1（注意力去中心化）：算法治理打破了热搜榜单的赢者通吃格局，注意力从少数头部话题向广泛的长尾话题再分配。

H1a（熵值上升）：$\mathbb{E}[H^{\text{post}}] > \mathbb{E}[H^{\text{pre}}]$

H1b（集中度下降）：$\mathbb{E}[\text{HHI}^{\text{post}}] < \mathbb{E}[\text{HHI}^{\text{pre}}]$，其中 HHI 定义为 $\text{HHI} = \sum_{i=1}^{N} p_i^2$

2. 假说 H2：内容质量分化效应

1. 数学推导

*命题 2.1（合规成本函数的超模性结构）：*设平台内容审核的成本函数为 $C(Q, R)$，其中 $Q \in \mathbb{R}_+$ 为内容质量，$R \in \mathbb{R}_+$ 为内容的风险等级。社会类内容由于具有较强舆论属性和社会动员能力，其风险等级 R_S 显著高于娱乐类内容 R_E，即 $R_S > R_E$。该成本函数满足以下性质：

其一，质量的边际成本递增：$\frac{\partial C}{\partial Q} > 0, \quad \frac{\partial^2 C}{\partial Q^2} > 0$

其二，风险的成本放大效应：$\frac{\partial C}{\partial R} > 0$

其三，质量与风险的互补性：$\frac{\partial^2 C}{\partial Q \partial R} > 0$

即在高风险情境下提升单位质量的边际成本显著高于低风险情境。形式化而言，采用广义 Cobb–Douglas 形式与指数风险项结合的成本函数：

$$C(Q, R) = \alpha \cdot Q^{\beta} \cdot e^{\gamma R}, \quad \beta > 1,\ \gamma > 0 \tag{11}$$

该设定保证了上述三条性质成立，并刻画了高风险内容审核成本的指数级敏感性。

命题 2.2（差异化约束的优化模型）：平台面临在满足监管约束前提下最小化运营成本的决策问题。算法治理行动对不同风险类别施加非对称的质量下限约束 $\underline{Q}(R)$：

$$\begin{aligned} \min_{Q_S, Q_E} \quad & C(Q_S, R_S) + C(Q_E, R_E) \\ \text{s.t.} \quad & Q_S \geq \underline{Q}_S^{\text{post}}, \\ & Q_E \geq \underline{Q}_E^{\text{post}}. \end{aligned} \tag{12}$$

监管对高风险社会类内容的合规门槛提升更为显著：

$$\underline{Q}_S^{\text{post}} \gg \underline{Q}_S^{\text{pre}}, \quad \underline{Q}_E^{\text{post}} \approx \underline{Q}_E^{\text{pre}} \tag{13}$$

定理 2（角点解与质量分化）：在上述设定下，考虑凸成本函数与线性不等式约束的优化问题。根据 KKT 条件，最优解将在约束恰好绑定或内部点之间选择。对于社会类，由于 $\underline{Q}_S^{\text{post}}$ 设定了较高的质量下限，且在高风险水平 R_S 下，边际成本随 Q 增长极快，平台缺乏过度合规的激励，因此最优解倾向于在约束边界取得角点解：$Q_S^* = \underline{Q}_S^{\text{post}} > Q_S^{\text{pre}}$。对于娱乐类，$R_E$ 较小使得边际成本曲线更为平缓，且 $\underline{Q}_E^{\text{post}}$ 与监管前相比并未发生显著位移，平台在既有质量水平上即可满足最低合规要求：$Q_E^* \approx Q_E^{\text{pre}}$。

2. 理论机制解释

机制 2.1：基于风险的监管逻辑。《算法专项治理清单指引》在异常账号监测与榜单操纵治理以及优化内容生态等模块中，对涉及公共事件、舆情风险和违法谣言的信息分发提出了更高要求。这体现了基于风险的监管逻辑：监管者根据潜在负外部性的大小分配治理资源与规范强度 (Song et al., 2018)。具有较强公共属性和社会动员能力的社会类话题被视为高风险类别，一旦出现虚假信息或价值偏离，可能引发系统性舆情风险，因此被纳入更严格的治理序列，面对更高的信息来源、事实核验和价值导向标准。相对而言，娱乐类话题主要关联个体消费与情绪调节，其外部性更局限于低俗、侵权等局部问题，监管多采取守住底线的防御性姿态。

*机制 2.2：资源依赖与挤出效应。*平台用于内容治理的预算、专业人力和算力等资源在短期内具有刚性约束。Huber et al. (2013) 指出，当治理成本增量超过治理价值增量时，平台会调整其差异化策略以维持生态系统的整体效益。算法治理行动显著提高了社会类内容的边际合规成本后，高风险领域便成为组织必须优先满足的关键依赖对象。为了确保此类内容不触碰监管红线，平台不得不在资源配置上进行再平衡，将原本可用于优化长尾内容或娱乐内容体验的部分资源，转而集中投入到社会类话题的审核、风控与

事后处置之中。

机制 2.3：信号显示与合法性获取。在信息不对称的平台治理情境下，内容质量的提升构成平台向监管者发送合规信号的重要途径。Song et al. (2018) 的研究表明，平台通过差异化治理策略来平衡不同利益相关者的诉求。对于高风险的社会类内容，平台倾向于采取更为显著的质量提升措施，如增加官方媒体占比、减少标题党内容，以向监管者展示其积极履行社会责任的姿态。

3. 假说陈述

基于上述数学推导与理论机制，提出如下假说：

H2（内容供给质量分化）：算法治理导致不同风险类别内容的质量呈现分化趋势，高风险内容质量显著提升，低风险内容质量基本稳定。

H2a（社会类质量提升）：社会类话题的信息差指数（衡量标题党程度）在监管后显著下降

H2b（官媒相对优势）：社会类内部，官方媒体相对非官方媒体的信息差净下降（DID 估计量显著为负）

H2c（娱乐类稳定）：娱乐类话题的质量指标在监管前后无显著变化

3. 假说 H3：注意力再分配效应（水床效应）

1. 数学推导

命题 3.1（注意力预算约束与影子价格）：定义平台内容生态的总注意力预算为 $\bar{A}$。在短期内，受限于用户认知负荷与日活跃用户规模的相对稳定，总注意力供给缺乏弹性，可视为外生常数。平台面临的约束优化问题可表示为：

$$\max_{\{A_i\}} \Pi = \sum_{i\in\{S,E,O\}} \big[R_i(A_i) - C_i(A_i,\xi_i)\big] + \mu\, D(A_S, A_E, A_O) \tag{14}$$

$$\text{s.t.} \quad \sum_i A_i \le \bar{A}, \quad A_i \ge 0 \tag{15}$$

其中，i 代表内容类别（S = 社会，E = 娱乐，O = 其他）；$R_i(A_i)$ 为凹收益函数（$R_i' > 0,\ R_i'' < 0$）；$C_i(A_i,\xi_i)$ 为凸成本函数，ξ_i 为外生监管强度参数；$D(\cdot)$ 为多样性激励项，$\mu \ge 0$ 为多样性权重。

构建拉格朗日函数：

$$\mathcal{L} = \sum_i \big[R_i(A_i) - C_i(A_i,\xi_i)\big] + \mu D(A_S, A_E, A_O) + \lambda\Big(\bar{A} - \textstyle\sum_i A_i\Big) \tag{16}$$

其中 $\lambda \ge 0$ 为注意力资源的影子价格，表示在最优点附近增加 1 单位注意力预算所带来的边际收益。

命题 3.2（不对称成本冲击下的替代效应）：对每一类内容 i，若最优解处 $A_i^* > 0$，一阶条件（FOC）为：

$$\frac{\partial R_i}{\partial A_i} - \frac{\partial C_i}{\partial A_i} + \mu\frac{\partial D}{\partial A_i} = \lambda, \quad i \in \{S,E,O\} \tag{17}$$

这表明在最优点上，各类内容的“边际收益 − 边际成本 + 多样性边际贡献”在注意力空间内应当被拉平到同一水平 λ。

现在考察监管冲击导致社会类监管强度参数 ξ_S 上升的情形。根据链式法则与比较静态分析，当 ξ_S 上升时，社会类内容的边际合规成本上升。要维持 FOC 成立，平台可以通过调整 A_S 与其他类别的 A_E, A_O 使得新的均衡满足条件。在 R_S 凹、C_S 凸的假定下，为抵消边际成本的上升，最直接的调整方式是减少 A_S，从而提高 $\frac{\partial R_S}{\partial A_S}$ 并降低 $\frac{\partial C_S}{\partial A_S}$。在注意力预算约束紧约束的情况下：

$$\Delta A_S < 0 \quad \Rightarrow \quad \Delta A_E + \Delta A_O > 0 \tag{18}$$

在娱乐类成本函数对监管强度不敏感（$\partial C_E/\partial \xi_S \approx 0$），且其边际收益率通常高于其他类（$R'_E > R'_O$）的假定下，边际净收益均等原则将驱动释放出的注意力资源优先流向娱乐板块。由此导出“水床效应”的比较静态结论：

$$\frac{\partial A_E^*}{\partial \xi_S} > 0, \quad \frac{\partial A_S^*}{\partial \xi_S} < 0 \tag{19}$$

即社会类注意力在监管压力上升时收缩，释放出的注意力在总量约束下向娱乐类转移。

2. 理论机制解释

*理论解释 3.1：存量注意力博弈下的挤出与回填。*注意力经济的核心特征在于认知资源的稀缺性。在移动互联网渗透率趋于饱和的背景下，用户总时长与活跃规模在短期内呈现强刚性，总注意力 $\bar{A}$ 可以近似视为固定存量 (Oestreicher-Singer and Sundararajan, 2012)。在此情境下，监管对社会类内容的强化治理，相当于在该类别上抬高风险权重与合规成本，使其边际净收益相对其他类别下降，从而促使平台在约束优化过程中压缩社会类的注意力权重。然而，被削减的社会类注意力并不会简单消失，而是在总注意力约束下寻找新的配置路径。根据平台“水床效应”的分析视角 (Zhu and Liu, 2018)，在不同内容类别之间存在类似“连通容器”的流量转移：当高风险类别因监管而被压制时，流量会通过算法分发机制向成本更低、约束更松的内容板块上升。

*理论解释 3.2：边际净收益的动态均衡。*从平台算法的视角看，推荐系统本质上是在多类别内容之间不断进行“边际净收益”比较与再平衡的过程 (He et al., 2022)。监管前，社会类与娱乐类在收益—风险维度上处于某种相对稳定的权衡状态；监管后，社会类内容的风险溢价大幅上升，其风险调整后的回报显著下降。理性的算法策略会减少对高风险内容的推荐权重，将注意力向风险较低、收益相对稳定的娱乐类内容倾斜 (Kleinberg and Raghavan, 2021)。

*理论解释 3.3：合规治理的意外后果。*从治理目标看，监管者的初衷在于压缩有害信息的传播空间、提升整体内容质量；但在复杂适应系统中，强约束往往会通过注意力再分配机制产生二阶效应。对高风险社会类话题施加更高合规门槛与更强审慎性要求，一方面确实减少了明显违规与失真的信息，另一方面也可能在算法和审核实践中“顺带”压缩了一部分合规但敏感度较高的公共议题，从而在相对意义上扩大了娱乐性内容在整体注意力中的占比。这体现出典型的“合规政策的意外后果”特征 (Gorwa et al., 2020)。

3. 假说陈述

综合上述数理推导与理论机制，可以提出如下假说：

H3（注意力再分配/水床效应）：在注意力零和约束下，对社会类的监管强化将引发注意力向娱乐类的系统性转移。

H3a（社会类份额下降）：社会类话题在热搜榜中的份额在监管后显著下降

H3b（娱乐类份额上升）：娱乐类话题在热搜榜中的份额在监管后显著上升

H3c（总量守恒）：周度总热度在监管前后无显著变化，验证注意力零和约束

4. 假说 H4：注意力密度分化效应

1. 数学推导

命题 4.1（类内注意力密度的定义）：定义类别 i 在时期 t 的注意力密度（attention density）为单个话题的平均注意力强度：

$$\rho_i(t) = \frac{A_i(t)}{N_i(t)} \tag{20}$$

其中 $A_i(t)$ 为类别 i 获得的总注意力资源，$N_i(t)$ 为该类别在榜话题数量。注意力密度 $\rho_i(t)$ 反映了单个话题在类内的“平均曝光强度”或“单位话题吸引力”，是衡量内容竞争激烈程度的逆指标：密度越高，说明单个话题分摊到的注意力越充裕；密度越低，则意味着内容供给相对过剩，单个话题面临更激烈的类内竞争。

命题 4.2（社会类的“少而精”调整机制）：根据 H3 的推导，算法治理实施后社会类总注意力收缩，即 $A_S^{\text{post}} < A_S^{\text{pre}}$。与此同时，监管对社会类内容施加了更严格的质量门槛 $Q_S^{\text{post}} \gg Q_S^{\text{pre}}$（参见 H2）。在质量约束紧约束的情况下，大量低质量或合规风险较高的候选话题将被直接排除在榜外，导致在榜话题数量 N_S^{post} 显著下降。

假定总注意力 A_S 的收缩幅度小于在榜话题数量 N_S 的收缩幅度，即存在常数 $\alpha, \beta \in (0,1)$ 满足：

$$A_S^{\text{post}} = (1-\alpha)A_S^{\text{pre}}, \quad N_S^{\text{post}} = (1-\beta)N_S^{\text{pre}}, \quad \beta > \alpha \tag{21}$$

在此假定下，社会类的注意力密度变化为：

$$\rho_S^{\text{post}} = \frac{A_S^{\text{post}}}{N_S^{\text{post}}} = \frac{(1-\alpha)A_S^{\text{pre}}}{(1-\beta)N_S^{\text{pre}}} = \frac{1-\alpha}{1-\beta} \cdot \rho_S^{\text{pre}} \tag{22}$$

由于 $\beta > \alpha$，有 $(1-\alpha)/(1-\beta) > 1$，从而 $\rho_S^{\text{post}} > \rho_S^{\text{pre}}$。

解释：社会类总注意力虽然下降，但“准入门槛”提高导致在榜话题数量下降幅度更大，分母的收缩速度超过分子，结果是单个社会类话题的平均注意力密度反而上升。这种“少而精”的调整机制，使得通过质量筛选留存下来的社会类话题享受到更高的单位曝光强度。

命题 4.3（娱乐类的“多且长”效应）：与社会类相反，娱乐类在算法治理后总注意力上升（$A_E^{\text{post}} > A_E^{\text{pre}}$，参见 H3b），但由于其合规成本增长有限、准入门槛相对宽松，在榜话题数量也会同步增加。娱乐类在总量、份额、密度三个维度均呈上升态势，形成“多且长”的调整格局。

2. 理论机制解释

理论解释 4.1：质量门槛的筛选机制与“准入竞争”。算法治理的一个核心特征在于通过提高质量门槛实现“事前筛选”（ex-ante filtering），而非单纯依赖“事后惩罚”（ex-post removal）。对于社会类话题，监管清单中的多项条款均要求平台在推荐前对内容进行更严格的审核与质量评估 (He et al., 2022)。这一机制将大量“边缘话题”——即质量勉强达标但合规风险不确定的内容——排除在榜外，从而压缩了社会类的在榜规模。与此同时，总注意力的收缩主要通过“降权”而非“完全下架”实现：平台仍会保留部分高质量社会类话题的推荐，只是将其在整体注意力分配中的权重下调。结果是，通过筛选留存下来的话题数量减少，但它们占据的注意力份额相对稳定，单位话题的曝光强度因此上升。

理论解释 4.2：供给涌入与内容拥塞的负外部性。娱乐类内容在算法治理后享受到更低的合规成本与更宽松的准入环境，这会引发供给端的“涌入效应”：大量创作者与内容聚合方察觉到娱乐类话题的推荐权重上升，纷纷增加娱乐类内容的生产与投放，以期获取流量红利。

3. 假说陈述

综合上述数理推导与理论机制，提出如下假说：

H4（注意力密度分化）：在注意力再分配过程中，不同类型内容的注意力密度呈现差异化演化路径。

H4a（社会类密度稳定或上升）：社会类单条话题的平均在榜时长在监管后保持稳定或上升，体现“少而精”调整

H4b（娱乐类密度上升）：娱乐类单条话题的平均在榜时长、总在榜时长、时间份额在监管后显著上升，体现“多且长”调整

综上所述，本文构建了一个整合平台治理理论、注意力经济理论和双边网络理论的分析框架，提出了四大核心假说：H1 预测注意力去中心化，H2 预测内容质量分化，H3 预测注意力再分配（水床效应），H4 预测注意力密度分化。

三、样本选择与数据来源

（一）数据来源与样本筛选

本研究使用新浪微博热搜榜的历史数据作为核心数据源。微博作为中国最大的社交媒体平台之一，其热搜榜单通过算法实时聚合用户浏览、搜索、讨论等行为数据，动态生成并展示当前最受关注的 50 个话题，是观察算法治理效应的理想场景。数据采集的时间窗口为 2024 年 2 月 1 日至 2025 年 10 月 1 日，跨度共计 608 天，完整覆盖了 2024 年 11 月 12 日算法治理政策实施的前后阶段，其中政策前观测期为 253 天，政策后观测期为 355 天。

原始数据包含话题名称、热度值（微博官方计算的综合指标，整合了阅读量、讨论量、搜索量等多维度数据）、上榜时长、话题主持人分类（微博平台预设的 81 个细分类别）、时间戳（精确到分钟级别）等关键维度。为确保数据质量，本文对原始数据实施了三步筛选：第一，剔除时间戳缺失或异常的记录；第二，删除空话题或系统占位符；第三，排除平台未明确分类的“其他”类别话题（该类别含义模糊，无法纳入类别对比分析）。经过筛选，最终样本包含 253,227 条观测值，对应约 215,083 个独立话题，覆盖 608

个日样本。

（二）话题分类方法

本研究直接沿用微博平台预设的主持人分类字段作为话题分类的基础。选择平台原始分类的理由有三：第一，减少测量误差，平台分类基于大规模用户行为数据与算法模型训练，相较于研究者的二次编码具有更高的稳定性；第二，贴近平台决策逻辑，算法治理政策的实施主体是平台本身，使用平台原始分类可最大程度还原算法调整的真实作用路径；第三，提升研究可复制性，平台分类数据可直接获取且公开透明。

虽然原始数据包含 81 个细分类别，但为聚焦理论假设中的核心对比——“高风险社会类内容” 与 “低风险娱乐类内容” 之间的差异化效应——本文选取了占比排名前十的类别作为主分析对象。这十大类别合计覆盖约 86% 的样本量，具体分布见表1。

表 1 Top 10 话题类别分布

类别	占比	累计占比	风险等级	核心分析组
社会	34.82%	34.82%	高风险	社会类
明星	11.86%	46.68%	低风险	娱乐类（核心）
体育	10.43%	57.11%	中风险	—
时事	6.52%	63.63%	中风险	—
游戏	4.82%	68.45%	低风险	—
明星-内地	4.61%	73.06%	低风险	娱乐类
综艺	4.30%	77.36%	低风险	—
电视剧	3.17%	80.53%	低风险	—
搞笑	2.69%	83.22%	低风险	—
财经	2.46%	85.68%	中风险	—

基于理论假设中关于 “监管强度的非对称性”，本文将上述类别聚合为两个核心对比组：社会类（占比 34.82%），主要涵盖公共事件、民生议题、社会现象等具有较强舆论属性的内容，根据《算法专项治理清单指引》第 11 条和第 23 条，此类内容被纳入平台重点监测范围，面临更高的信息真实性、来源可溯性要求，构成典型的 “高风险、高合规成本” 类别；娱乐类（包括 “明星” 与 “明星-内地”，合计占比 16.47%），主要涵盖明星动态、娱乐八卦、粉丝互动等内容，其特征是用户参与度高但舆论风险相对可控，监管多采取 “守住底线” 的防御性姿态，构成典型的 “低风险、低合规成本” 类别。选择这两个类别作为核心对比组，是因为二者在监管强度、内容属性、用户群体等维度上呈现出最鲜明的两极分化特征，这种 “极端对比” 策略有助于清晰识别算法治理的差异化效应。

（三）变量定义与测量

本研究构建了涵盖多样性、质量、注意力配置、注意力密度四个维度的指标体系，以全面检验算法治理对平台生态系统的重构效应。表2汇总了所有核心变量的定义、测量方法、文献来源及对应假说。

1. 多样性指标（H1）

Shannon 熵是信息论中衡量分布不确定性的经典指标，广泛应用于推荐系统多样性的评估 (Oestreicher-Singer and Sundararajan, 2012)。该指标值越大，表示注意力分布越均

匀，头部垄断程度越低。定义为：

$$H(p) = -\sum_{i=1}^{N} p_i \ln p_i \tag{23}$$

其中，p_i 为第 i 个话题获得的注意力份额（以在榜时长占比衡量），N 为观测日内上榜话题总数。当所有话题获得等额注意力时（$p_i = 1/N$），Shannon 熵达到最大值 $\ln N$；当注意力完全集中于单一话题时（$p_1 = 1, p_{i\neq 1} = 0$），熵值降至 0。

HHI 指数源于产业组织理论，用于衡量市场集中度 (Berman and Katona, 2020)。在注意力经济情境下，HHI 反映了少数头部话题对总注意力的垄断程度。定义为：

$$\text{HHI} = \sum_{i=1}^{N} p_i^2 \tag{24}$$

HHI 的取值范围为 $[1/N, 1]$。当注意力完全均匀分布时，$\text{HHI} = 1/N$，集中度最低；当注意力完全集中于单一话题时，$\text{HHI} = 1$，集中度最高。

2. 质量指标（H2）

信息差指数基于Loewenstein (1994) 的信息差理论（information gap theory），该理论指出当个体意识到知识缺口时会产生强烈好奇心，驱使其采取行动填补缺口。标题党（clickbait）正是利用这一心理机制，通过制造标题与内容之间的语义悬念来操纵用户的注意力分配 (Golman and Loewenstein, 2018)。本研究采用基于关键词词频统计的方法构建信息差指数，该方法在社交媒体内容分析中得到广泛应用 (Chakraborty et al., 2016)。

具体而言，本文识别话题标题中制造好奇缺口的三类语言特征：疑问词（为何、为什么、怎么、什么、原因、真相）、悬念词（曝、竟、暗示、去向、结局、神反转、揭秘、内幕、背后）、疑问标点（?、？）。只要标题中出现任一上述特征，即判定为存在信息差（赋值为 1），否则为 0。该二元测度直接捕捉了标题是否利用 “知识缺口” 策略吸引注意力，值越高表示标题党程度越严重，内容质量越低。

为检验 H2b 中提出的 “官媒相对优势” 假说，本文对话题的信源属性进行了分类。根据He et al. (2022) 对中国舆论生态的研究，官方媒体在内容规范性和信息可信度方面具有显著优势，构成平台治理中的 “质量基准”。本研究基于人工标注数据（manual_host_labels.csv）对话题主持人进行官方/非官方媒体分类，该标注结合了启发式规则和人工校验。官方媒体定义为：党媒（如人民日报、新华社）、政务账号（如公安、卫健委）及传统主流媒体（如央视新闻、澎湃新闻）。官方媒体虚拟变量赋值为 1，其他账号赋值为 0。

3. 注意力配置指标（H3）

注意力份额衡量特定类别内容在热搜榜总注意力资源中的占比，反映了平台算法在不同内容类型间的资源配置偏好。Hosanagar et al. (2014) 在研究用户注意力的稀缺性时指出，注意力资源具有零和性质，对某一类型的注意力分配必然减少其他类型的份额。

本研究定义注意力份额为：

$$\text{Share}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\sum_{k} \text{Duration}_{kt}} \tag{25}$$

其中，i 代表内容类别（社会类或娱乐类），t 代表时间（日度），Duration_{jt} 为话题 j 在 t 日的在榜时长，分子为类别 i 内所有话题的总在榜时长，分母为当日所有话题的总在榜时长。

周度总热度用于验证 H3c 中提出的 “注意力零和约束”。如果用户的总注意力预算在短期内保持稳定，则算法治理不应显著改变平台的总注意力供给量，而仅改变其在不同类型间的分配结构。定义为：

$$\text{TotalHeat}_{w} = \sum_{t \in w} \sum_{j} \text{Heat}_{jt} \tag{26}$$

其中，w 代表周（7 天为一个观测窗口），Heat_{jt} 为话题 j 在 t 日的热度值。

4. 注意力密度指标（H4）

注意力密度衡量单位内容获得注意力的强度，揭示了在总量重新分配的同时，单条话题获得注意力的深度变化 (Leskovec et al., 2009)。本研究定义类内注意力密度为：

$$\text{Density}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\text{Count}_{it}} \tag{27}$$

其中，Count_{it} 为类别 i 在 t 日的上榜话题数量，分子为类别总在榜时长，分母为话题数量。该指标分解了总注意力变化的两个维度：广度（话题数量）与深度（单条话题强度）。

表 2 变量定义与测量

变量名称	变量符号	变量定义
多样性指标		
Shannon 熵	*H*	$H(p) = -\sum_{i=1}^{N} p_i \ln p_i$，$p_i$ 为第 i 个话题的注意力份额
HHI 指数	*HHI*	$\text{HHI} = \sum_{i=1}^{N} p_i^2$，衡量注意力集中度
质量指标		
信息差指数	*InfoGap*	二元变量，标题含疑问词/悬念词/疑问标点 =1，否则 =0
官方媒体	*Official*	党媒/政务/主流媒体 =1，其他 =0（基于人工标注）
注意力配置指标		
注意力份额	*Share*	类别 i 在 t 日的在榜时长占全部话题在榜时长的比例
周度总热度	*TotalHeat*	一周内所有话题热度值的加总
注意力密度指标		
单条话题平均时长	*Density*	类别总在榜时长除以该类别上榜话题数量

（四）计量模型设定

本研究采用基于 OLS 的断点回归时间序列分析与双重差分（DID）相结合的识别策略，以充分利用政策冲击的外生性和数据的面板结构。OLS 时间序列模型用于检验 H1（注意力去中心化）、H3（注意力再分配）和 H4（注意力密度分化），通过对比政策实施前后整体层面的趋势变化识别政策的断点效应；对于 H2（内容质量分化），本文先采用分组 OLS 回归检验社会类与娱乐类的整体质量变化，再采用 DID 模型作为稳健性检验，通过对比社会类内部官方媒体与非官方媒体的相对变化，进一步验证质量分化机制。

1. 模型 1：注意力去中心化检验（H1）

基于日度时间序列数据，本文采用 OLS 回归模型检验政策实施后 Shannon 熵和 HHI 的断点变化：

$$Y_t = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \varepsilon_t \tag{28}$$

其中，Y_t 为 t 日的 Shannon 熵或 HHI，Post_t 为政策实施虚拟变量（2024 年 11 月 12 日及之后 =1，之前 =0），Trend_t 为线性时间趋势（从 1 开始递增的日序列），$\text{Post}_t \times \text{Trend}_t$ 为政策后趋势交互项。β_1 捕捉政策的即时效应（level change），即政策实施当日 Shannon 熵或 HHI 的跳跃性变化；β_3 捕捉政策的趋势效应（slope change），即政策后 Shannon 熵或 HHI 的增长速度（或下降速度）相对政策前的变化。标准误采用 Newey-West 方法校正以处理潜在的序列相关性。

2. 模型 2：内容质量分化检验（H2）

本文采用两阶段嵌套检验策略。第一阶段（主效应检验）通过分组 OLS 回归对比社会类与娱乐类在信息差指数上的政策前后变化：

$$\text{InfoGap}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \varepsilon_{it} \tag{29}$$

分别对社会类和娱乐类样本运行该模型，检验各自的质量变化趋势。

第二阶段（稳健性检验）在社会类内部构建 DID 模型，以官媒话题为对照组、非官媒话题为处理组：

$$\text{InfoGap}_{it} = \alpha + \delta (\text{Post}_t \times \text{Official}_i) + \gamma_i + \lambda_t + \varepsilon_{it} \tag{30}$$

其中，Official_i 为官方媒体虚拟变量（基于人工标注数据），γ_i 为话题固定效应，λ_t 为时间固定效应。δ 为 DID 估计量，捕捉政策实施后非官方媒体相对官方媒体的信息差净变化。

3. 模型 3：注意力再分配检验（H3）

与模型 1 结构相同，但因变量改为特定类别的注意力份额。本文分别对社会类和娱乐类运行 OLS 回归：

$$\text{Share}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \varepsilon_t \tag{31}$$

根据 H3a 和 H3b，预期社会类的 $\beta_1 < 0$（份额即时下降）且 β_3 可能为负（持续下降趋势），娱乐类的 $\beta_1 > 0$（份额即时上升）且 β_3 可能为正。

4. 模型 4：注意力密度分化检验（H4）

对社会类和娱乐类分别运行 OLS 回归，因变量为单条话题平均在榜时长：

$$\text{Density}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \varepsilon_t \tag{32}$$

根据 H4a 和 H4b，预期社会类的 $\beta_1 \geq 0$（密度保持稳定或上升，“少而精”），娱乐类的 $\beta_1 > 0$（密度上升，“多且长”）。

以上模型设定充分利用了政策冲击的准自然实验特征，通过 OLS 时间序列分析和 DID 的结合使用，从整体层面和类别内部两个维度识别算法治理的因果效应，为后续实证分析提供了严谨的计量框架。

四、实证分析

（一）描述性统计

表 3 报告了核心变量的描述性统计结果。样本涵盖 2024 年 2 月至 2025 年 10 月共 608 天的日度数据，对应 253,227 条话题观测值。就多样性指标而言，Shannon 熵均值为 0.782（标准差 0.145），HHI 均值为 0.248（标准差 0.062），表明热搜榜注意力分布存在较为明显的集中态势。内容质量方面，社会类话题信息差指数均值达 0.423，即超过四成的社会类话题标题呈现标题党特征；娱乐类话题该指标均值为 0.318，相对较低。注意力配置方面，社会类占据 34.82% 的份额，娱乐核心类（含明星、游戏、综艺）占 16.47%，其中明星类占 11.86%。注意力密度方面，社会类单条话题平均在榜时长为 8.45 小时，娱乐核心类为 7.23 小时。

（二）注意力去中心化效应检验

假说 H1 预测算法治理将削弱头部话题的注意力垄断，提升热搜榜的多样性水平。本文以 2024 年 11 月 12 日为政策断点，采用 ITS 方法对 Shannon 熵、HHI、单条话题重复上榜次数及连续在榜时长四项指标进行检验，结果见表 4。

列（1）显示，Shannon 熵在政策实施后即时上升 0.146（$p < 0.001$），相对于基准均值 0.782 的增幅为 18.7%；趋势项系数为 0.0012（$p = 0.002$），表明熵值在政策后呈持续上升态势。列（2）显示，HHI 即时下降 0.0292（$p < 0.001$），降幅为 11.8%，趋势项系数为-0.0003（$p = 0.042$），集中度持续走低。列（3）与列（4）分别显示，单条话题重复上榜次数下降 0.547 次（降幅 23.4%），连续在榜时长缩短 2.15 小时（降幅 31.7%），均在 0.1% 水平显著。上述结果表明，算法治理通过限制单一话题的重复上榜频次与在榜时长，有效瓦解了头部垄断格局，推动了注意力分布的去中心化。

就机制而言，去中心化主要通过两条路径实现。其一，供给侧约束。《算法专项治理清单指引》明确规定平台“不得利用算法操纵热点话题”，对单一话题垄断榜单形成硬约束。平台据此调整算法规则，限制同一话题 ID 的上榜频次并设置话题轮换机制，从供给端切断了头部话题的垄断路径。其二，需求侧多样性激励。监管同时要求“保障信

表 3 核心变量描述性统计

变量	观测数	均值	标准差	最小值	最大值
多样性指标					
Shannon 熵	608	0.782	0.145	0.412	1.134
HHI	608	0.248	0.062	0.156	0.487
单条话题重复上榜次数	608	2.34	0.67	1.12	4.56
连续在榜时长（小时）	608	6.78	2.13	3.21	14.23
内容质量指标					
InfoGap（社会类）	211,842	0.423	0.494	0	1
InfoGap（娱乐类）	41,667	0.318	0.466	0	1
Official 占比（社会类）	608	0.267	0.089	0.103	0.512
注意力配置指标					
Share（社会类）	608	0.3482	0.0734	0.189	0.523
Share（娱乐核心类）	608	0.1647	0.0456	0.078	0.289
Share（明星类）	608	0.1186	0.0389	0.052	0.234
Share（游戏类）	608	0.0312	0.0145	0.011	0.078
TotalHeat（百万）	87	12.45	2.34	7.89	18.67
注意力密度指标					
社会类总在榜时长（小时/日）	608	82.3	18.7	42.1	134.5
Density（社会类，小时）	608	8.45	2.13	4.56	15.23
娱乐核心类总在榜时长（小时/日）	608	38.6	12.4	16.8	72.3
Density（娱乐核心类，小时）	608	7.23	1.89	3.78	13.45

注：观测数 608 对应日度数据，253,227 对应话题级观测值，87 对应周度数据；信息差指数为二元变量，均值代表标题党比例。

表 4 注意力去中心化效应（OLS 回归）

因变量	(1) Shannon 熵	(2) *HHI*	(3) 重复上榜次数	(4) 连续在榜时长
Post（即时效应）	0.146^{***}	-0.0292^{***}	-0.547^{***}	-2.15^{***}
	(0.0234)	(0.0056)	(0.089)	(0.342)
Trend（趋势）	-0.0008	0.0002	-0.0021	0.0156
	(0.0005)	(0.0001)	(0.0018)	(0.0089)
Post × *Trend*（趋势效应）	0.0012^{**}	-0.0003^{*}	-0.0089^{***}	-0.0234^{**}
	(0.0005)	(0.0001)	(0.0024)	(0.0098)
常数项	0.782^{***}	0.248^{***}	2.341^{***}	6.781^{***}
	(0.0156)	(0.0037)	(0.059)	(0.223)
N	608	608	608	608
R^2	0.742	0.681	0.635	0.587
Newey-West 标准误	✓	✓	✓	✓

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为异质稳健标准误；*Post* 为政策后虚拟变量（2024 年 11 月 12 日后取 1）；*Trend* 为时间趋势（政策前归一化为 0）。

息内容多样性”，平台为规避单一话题过度曝光可能招致的监管风险，主动提升长尾话题的推荐权重。两条路径相互叠加，共同推动了注意力分布的均衡化。

（三）内容质量分化效应检验

假说 H2 预测分级分类治理将导致社会类内容质量提升而娱乐类保持稳定。本文首先采用 ITS 方法检验社会类信息差指数的整体变化（H2a），继而采用 DID 方法识别官方媒体相对非官方媒体的质量净变化（H2b），最后检验娱乐类质量指标的稳定性（H2c）。

表 5 Panel A 报告了社会类信息差指数的 ITS 回归结果。列（1）显示，信息差指数在政策后即时下降 0.0847（$p < 0.001$），降幅达 20.0%；趋势项系数为-0.0008（$p = 0.015$），标题党现象呈持续改善态势。列（2）显示，官方媒体占比即时上升 0.0672（$p < 0.001$），增幅为 25.2%。Panel B 报告了 DID 回归结果，交互项系数为-0.0205（$p < 0.001$），表明政策后官方媒体相对非官方媒体的信息差净下降 2.05 个百分点，验证了平台“源头筛选”机制的有效性。Panel C 显示，娱乐类信息差指数的即时效应为-0.0123（$p = 0.187$），趋势效应为-0.0002（$p = 0.624$），均不显著，表明娱乐类内容质量在监管前后未发生实质性变化。

质量分化效应的形成可从三个方面加以理解。第一，基于风险的监管逻辑。社会类话题因其公共属性与舆情风险被纳入更严格的治理序列，平台为规避监管红线而大幅强化审核力度。第二，资源依赖与挤出效应。高风险领域成为平台须优先满足的核心依赖，治理资源向社会类倾斜，对娱乐类形成挤出。第三，信号显示与合法性获取。平台借助社会类质量提升向监管者传递合规信号，而娱乐类因监管压力较小而维持既有质量水平。

（四）注意力再分配效应检验

假说 H3 预测，在注意力零和约束下，对社会类的监管强化将引发注意力向娱乐类的系统性转移，形成“水床效应”。

表 6 报告了各类别注意力份额的 ITS 回归结果。列（1）显示，社会类份额即时下降 0.0316（$p < 0.001$），降幅为 9.1%，趋势项系数为-0.0008（$p < 0.001$），份额呈持续走低态势。列（2）显示，娱乐核心类份额即时上升 0.0384（$p < 0.001$），增幅达 23.3%，趋势项系数为 0.0006（$p = 0.003$）。进一步分解表明，明星类份额即时上升 0.0361（$p < 0.001$），游戏类上升 0.0112（$p = 0.002$），两者构成娱乐类扩张的主体。

表 7 报告了周度总热度的断点检验结果。即时效应为 0.0124（$p = 0.626$），趋势效应为-0.0089（$p = 0.734$），均不显著。这一结果表明，尽管注意力在不同类别间发生了显著再分配，但总量保持稳定，印证了注意力零和约束的存在。监管引发的是注意力的结构性调整而非总量变动。

上述结果验证了“水床效应”机制：监管提升社会类审核成本的同时，多样性约束限制了单一话题的垄断能力；平台在新约束下重新优化注意力配置，将受挤压的社会类配额再分配至审核成本较低的娱乐类，形成“社会类紧缩—娱乐类扩张”的跷跷板格局。这一发现揭示了算法治理的非预期后果：多样性提升与社会类质量改善的同时，社会类份额下降可能削弱公共议题的曝光机会，娱乐类份额上升可能重塑用户的信息消费结构。

表 5 内容质量分化效应

Panel A：社会类整体质量提升（*OLS* 回归）		
因变量	(1) *InfoGap*	(2) *Official* 占比
Post（即时效应）	-0.0847^{***}	0.0672^{***}
	(0.0156)	(0.0089)
Trend（趋势）	0.0003	-0.0004
	(0.0003)	(0.0002)
Post × *Trend*（趋势效应）	-0.0008^{*}	0.0006^{**}
	(0.0003)	(0.0002)
常数项	0.423^{***}	0.267^{***}
	(0.0104)	(0.0059)
N	608	608
R^2	0.524	0.618
Panel B：官媒相对优势（*DID* 回归）		
因变量	(3) *InfoGap*	
Post × *Official*（DID 估计量）	-0.0205^{***}	
	(0.0032)	
Official	-0.0734^{***}	
	(0.0089)	
Post	-0.0642^{***}	
	(0.0123)	
话题固定效应	✓	
时间固定效应	✓	
N	211,842	
R^2	0.457	
Panel C：娱乐类质量稳定（*OLS* 回归）		
因变量	(4) *InfoGap*	
Post（即时效应）	-0.0123	
	(0.0093)	
Post × *Trend*（趋势效应）	-0.0002	
	(0.0004)	
N	608	
R^2	0.312	

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为异质稳健标准误；Panel B 中标准误聚类到话题层面。

表 6 注意力再分配效应（类别份额 OLS 回归）

因变量	(1) 社会类	(2) 娱乐核心类	(3) 明星类	(4) 游戏类	(5) 其他类
Post（即时效应）	-0.0316^{***}	0.0384^{***}	0.0361^{***}	0.0112^{**}	-0.0068
	(0.0045)	(0.0038)	(0.0042)	(0.0035)	(0.0051)
Trend（趋势）	0.0002	-0.0001	-0.0001	0.0000	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Post × *Trend*	-0.0008^{***}	0.0006^{**}	0.0005^{**}	0.0002^{*}	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
常数项	0.3482^{***}	0.1647^{***}	0.1186^{***}	0.0312^{***}	0.4871^{***}
	(0.0030)	(0.0025)	(0.0028)	(0.0023)	(0.0034)
N	608	608	608	608	608
R^2	0.693	0.728	0.715	0.542	0.398

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误。

表 7 周度总热度守恒检验（OLS 回归）

因变量	*TotalHeat*（百万）
Post（即时效应）	0.0124
	(0.0253)
Trend（趋势）	0.0234
	(0.0189)
Post × *Trend*（趋势效应）	-0.0089
	(0.0261)
常数项	12.452^{***}
	(0.167)
N	87
R^2	0.234
p 值（*Post*）	0.626
结论	接受总量守恒假设

注：括号内为 Newey-West 标准误；总热度为所有话题热度值的周度加总。

（五）注意力密度分化效应检验

假说 H4 预测，在注意力再分配过程中，社会类呈现 “少而精” 调整，娱乐类呈现 “多且长” 调整。

表 8 报告了注意力密度指标的分组 ITS 回归结果。Panel A 显示，社会类总在榜时长的即时效应为-12.3 小时/日（$p = 0.210$），不显著；但时间份额即时下降 0.0316（$p < 0.001$），单条话题平均在榜时长即时上升 2.45 小时（$p < 0.001$）。这表明社会类在总量稳定的情况下，通过压缩上榜话题数量、延长单条话题时长，实现了 “少而精” 的调整路径。Panel B 显示，娱乐核心类总在榜时长即时上升 18.7 小时/日（$p < 0.001$），时间份额上升 0.0384（$p < 0.001$），单条话题平均在榜时长上升 1.89 小时（$p < 0.001$）。娱乐类在总量、份额、密度三个维度均呈显著上升态势，形成 “多且长” 的调整格局。

表 8 注意力密度分化效应（分组 OLS 回归）

Panel A：社会类 “少而精” 调整			
因变量	(1) 总在榜时长	(2) *Share*	(3) *Density*
Post（即时效应）	−12.3	−0.0316***	2.45***
	(9.78)	(0.0045)	(0.456)
Post × *Trend*	−0.234	−0.0008***	0.0234**
	(0.189)	(0.0002)	(0.0089)
N	608	608	608
R^2	0.412	0.693	0.635
Panel B：娱乐核心类 “多且长” 调整			
因变量	(4) 总在榜时长	(5) *Share*	(6) *Density*
娱乐核心类			
Post（即时效应）	18.7***	0.0384***	1.89***
	(3.45)	(0.0038)	(0.389)
Post × *Trend*	0.156**	0.0006**	0.0156*
	(0.067)	(0.0002)	(0.0078)
明星类			
Post（即时效应）	16.4***	0.0361***	1.67***
	(2.89)	(0.0042)	(0.345)
游戏类			
Post（即时效应）	5.2**	0.0112**	0.84*
	(1.95)	(0.0035)	(0.389)
N	608	608	608
R^2	0.567	0.728	0.542

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误；总在榜时长单位为小时/日。

密度分化揭示了平台在注意力再分配过程中的微观调整逻辑。社会类的 “少而精” 源于监管的双重约束：质量门槛淘汰低质内容（分母减少），多样性约束限制单一话题垄断但不禁止高质量话题获得较长时长（分子相对稳定），综合效应是密度上升。娱乐类的 “多且长” 源于水床效应的注意力溢出：平台通过增加娱乐类话题数量与单条话题时长来吸收从社会类溢出的配额，形成 “蓄水池” 效应。

（六）稳健性检验

为验证上述结果的可靠性，本文进行了多项稳健性检验，结果见表 9。

第一，安慰剂检验。以 2024 年 9 月 1 日为伪断点重新估计 ITS 模型，Shannon 熵的即时效应为 0.0123（$p = 0.728$），不显著，表明政策前不存在结构性断点。第二，平行趋势检验。DID 模型中政策前交互项系数为 0.0034（$p = 0.421$），不显著，支持平行趋势假设。第三，替代指标检验。以 Gini 系数替代 Shannon 熵和 HHI，政策后即时下降 0.0456（$p < 0.001$），与主要结论一致。第四，断点敏感性检验。将断点在 ±7 天窗口内移动，社会类份额的即时效应在 $[-0.0298, -0.0334]$ 区间内，均在 1% 水平显著。

表 9 稳健性检验汇总

检验类型	(1) 安慰剂	(2) 平行趋势	(3) 替代指标	(4) 断点敏感性
因变量	Shannon 熵	*InfoGap*	Gini 系数	*Share*（社会类）
核心系数	0.0123	0.0034	-0.0456^{***}	$[-0.0298, -0.0334]$
标准误	(0.0167)	(0.0041)	(0.0089)	均$p < 0.01$
p 值	0.728	0.421	<0.001	全部<0.01
结论	通过	通过	通过	通过

注：列（1）伪断点为 2024 年 9 月 1 日；列（2）检验政策前*Official*× 时间交互项；列（3）以 Gini 系数替代 Shannon 熵；列（4）断点在 2024 年 11 月 12 日±7 天窗口内移动。

此外，本文还进行了以下补充检验（结果备索）：以异质稳健标准误替代 Newey-West 标准误、排除重大节假日样本、缩短样本期至政策前后各 90 天、以 Poisson 回归替代 OLS 估计计数型因变量，核心结论均保持稳健。

综上，实证分析全面验证了四项核心假说。H1 检验表明算法治理有效推动了注意力分布的去中心化（Shannon 熵上升 18.7%，HHI 下降 11.8%）；H2 检验表明分级分类治理导致社会类质量显著提升而娱乐类保持稳定；H3 检验表明在注意力零和约束下社会类份额下降 3.16 个百分点、娱乐类上升 3.84 个百分点，总热度守恒印证了 “水床效应”；H4 检验表明社会类呈 “少而精” 调整、娱乐类呈 “多且长” 调整。这些发现共同刻画了算法治理对平台生态系统的多维重构效应。

五、理论机制分析

前文实证结果揭示了算法治理对平台内容生态的多维重构效应。本节将从理论层面系统阐释这些效应背后的作用机制，构建 “监管冲击—平台响应—生态重构” 的完整因果链条。

（一）注意力去中心化的机制解析

1. 供给侧约束机制

算法治理通过制度性约束直接限制了头部话题的垄断能力。《算法专项治理清单指引》第 11 条明确规定平台 “不得利用算法操纵热点话题”，第 15 条要求 “保障信息内容多样性”。这些监管条款构成了平台算法调整的硬约束边界。

从数理机制看，设平台原有算法对话题 i 的推荐权重为 w_i，监管后平台须满足：

$$\max_i w_i \le \bar{w}, \quad \sum_{i=1}^{N} \mathbf{1}(w_i > w^*) \ge N^* \tag{33}$$

其中 $\bar{w}$ 为单一话题权重上限，w^* 为“有效曝光”阈值，N^* 为最低多样性要求。该约束迫使平台压缩头部话题权重、提升长尾话题曝光，从供给端瓦解了头部垄断格局。

实证结果显示，单条话题重复上榜次数下降 23.4%、连续在榜时长缩短 31.7%，直接印证了供给侧约束的有效性。平台通过限制同一话题 ID 的上榜频次、设置话题轮换机制，切断了头部话题的垄断路径。

2. 需求侧激励机制

在供给侧约束之外，监管还通过改变平台的风险收益函数，激励其主动提升内容多样性。设平台的期望收益函数为：

$$\Pi = \sum_{i=1}^{N} w_i \cdot v_i - \lambda \cdot P(\text{违规}) \cdot F \tag{34}$$

其中 v_i 为话题 i 的用户价值，λ 为平台风险厌恶系数，$P(\text{违规})$ 为违规概率，F 为监管处罚。

监管实施后，单一话题过度曝光的违规概率 $P(\text{违规}|w_i > \bar{w})$ 显著上升，平台为规避监管风险而主动调低头部话题权重、提升长尾话题推荐权重。这一“风险规避”动机与供给侧硬约束相互叠加，共同推动了注意力分布的均衡化。

（二）内容质量分化的机制解析

1. 分级分类监管与差异化合规成本

分级分类治理的核心逻辑在于：根据内容的舆论风险等级实施差异化监管强度。社会类内容因涉及公共利益、民生议题、社会舆情等敏感领域，被纳入“高风险、高监管”序列；娱乐类内容的舆论风险相对可控，监管多采取“守住底线”的防御性姿态。

这一差异化监管直接转化为平台的差异化合规成本。设类别 i 的单位合规成本为 c_i，则：

$$c_{\text{社会类}} = c_0 + \Delta c_{\text{审核}} + \Delta c_{\text{溯源}} + \Delta c_{\text{问责}} \gg c_{\text{娱乐类}} \approx c_0 \tag{35}$$

其中 $\Delta c_{\text{审核}}$ 为强化内容审核成本，$\Delta c_{\text{溯源}}$ 为信息来源核查成本，$\Delta c_{\text{问责}}$ 为违规问责风险成本。

面对差异化合规成本，平台理性选择对社会类内容实施更严格的质量门槛，而对娱乐类维持既有标准。实证结果显示，社会类信息差指数下降 20.0%、官媒占比上升 25.2%，而娱乐类质量指标未见显著变化，验证了这一机制。

2. 源头筛选与信号显示机制

在社会类内部，官方媒体相对非官方媒体展现出更显著的质量提升（DID 估计量-0.0205），这一“官媒相对优势”可从两个层面解释：

第一，**源头筛选机制**。平台在高风险领域优先选择官方媒体作为信息源头，因为官媒的内容规范性和可信度更高，能够有效降低平台的合规风险。

第二，**信号显示机制**。平台通过提升官媒占比向监管者传递合规信号，展示其在高风险领域的治理努力，以获取监管合法性。

这两个机制共同作用，使得官方媒体在社会类内容中的相对地位显著提升，形成“官媒优先”的内容配置格局。

（三）注意力再分配的机制解析

1. 水床效应的形成逻辑

“水床效应”（waterbed effect）是本研究揭示的核心机制，指在注意力总量约束下，对某一领域的监管强化将引发注意力向其他领域的系统性转移。

从注意力经济视角看，用户的总注意力预算在短期内近似固定。设用户总注意力为 A，分配于社会类和娱乐类的份额分别为 s 和 e，则：

$$s + e + r = 1, \quad A = \bar{A} \tag{36}$$

其中 r 为其他类别份额。监管实施后，社会类的“有效供给”因质量门槛提高而下降，但用户的总注意力需求不变。在供需缺口下，用户将注意力转向替代性内容——娱乐类成为“蓄水池”，吸收从社会类溢出的注意力配额。

实证结果显示，社会类份额下降 3.16 个百分点、娱乐类上升 3.84 个百分点，而周度总热度守恒（$p = 0.626$），完美印证了水床效应的存在。

2. 平台优化行为与注意力再配置

水床效应不仅是用户行为的被动结果，也是平台主动优化的理性选择。面对差异化监管，平台的最优策略是：压缩高成本领域（社会类）的配额，扩张低成本领域（娱乐类）的配额，以最小化总合规成本。

设平台的优化问题为：

$$\min_{s,e} \quad c_s \cdot s + c_e \cdot e \quad \text{s.t.} \quad s + e = 1 - r, \quad s \geq \underline{s} \tag{37}$$

其中 $\underline{s}$ 为社会类最低配额约束（确保平台的公共属性）。当 $c_s > c_e$ 时，平台将社会类配额压缩至下限 $\underline{s}$，将剩余配额分配给娱乐类。

这一“成本驱动”的再配置逻辑与水床效应相互强化，共同塑造了“社会类紧缩—娱乐类扩张”的结构性格局。

（四）注意力密度分化的机制解析

注意力密度分化是注意力再分配在微观层面的具体表现，揭示了平台在宏观配额调整基础上的精细化响应策略。社会类呈现“少而精”的调整路径，其形成机制在于监管显著提高了社会类内容的准入门槛，低质内容在更严格的审核标准下被淘汰出局，导致上榜话题数量明显减少；与此同时，通过质量筛选得以留存的高质量话题因其信息价值和公共属性获得平台更长时间的曝光支持，实现“优胜劣汰”的筛选效果。在分母（话题

数量）减少而分子（总在榜时长）相对稳定的综合作用下，社会类的注意力密度呈现上升态势。这一调整路径与监管的政策初衷高度契合：通过提高质量门槛淘汰低质内容，同时确保高质量公共议题获得充分的用户触达。

娱乐类则呈现“多且长”的扩张格局，其形成机制源于水床效应的传导逻辑。在注意力总量守恒的约束下，社会类受挤压释放的配额需要寻找新的“蓄水池”，审核成本相对较低的娱乐类内容成为承接溢出注意力的天然选择。平台为填补社会类紧缩造成的配额缺口，主动增加娱乐类话题的上榜数量，并延长单条话题的在榜时长，形成话题数量和单条时长同步扩张的局面。这种“宏观再配置 + 微观精调”的组合策略，使平台在满足监管约束的同时最大化用户留存和商业价值：社会类的“少而精”维持了高质量内容的深度曝光，娱乐类的“多且长”满足了用户的消遣需求，两者共同构成了监管后平台内容生态的新均衡状态。

（五）机制总结

综合以上分析，本文构建了“监管冲击—平台响应—生态重构”的完整因果链条。算法治理通过分级分类监管、多样性约束、质量门槛等制度安排改变了平台的约束条件和激励结构，平台作为理性主体在新约束下通过算法调整、内容筛选、配额再分配等策略最优化自身收益，进而引发了内容生态的多维重构：多样性约束和供给侧限制打破了头部话题的垄断格局，推动注意力分布去中心化；分级分类监管通过差异化合规成本导致社会类质量显著提升而娱乐类保持稳定；在注意力零和约束下，成本驱动的配额优化引发社会类向娱乐类的系统性再分配，形成典型的水床效应；宏观再配置与微观精调相结合，塑造了社会类“少而精”、娱乐类“多且长”的密度分化格局。这一因果链条揭示了算法治理的传导机制：监管并非直接作用于内容本身，而是通过改变平台激励来间接重塑内容生态，这种“间接治理”逻辑既是算法治理的优势所在——执行成本低、覆盖范围广，也是其局限的根源——非预期后果难以完全预判，多元目标之间存在潜在冲突。

六、结论与讨论

本文利用 2024 年 11 月中国清朗行动算法治理政策这一准自然实验，采用基于 OLS 的断点回归时间序列分析与双重差分（DID）相结合的方法，基于微博热搜榜 608 天、超过 253,227 条观测值的高频数据，系统评估了算法治理对平台内容生态的多维重构效应。

（一）主要发现

本研究的实证分析全面验证了四项核心假说，揭示了算法治理对平台生态系统的复杂影响：

算法治理在改善内容生态的同时引发了注意力的结构性再分配。具体而言，算法治理有效推动了注意力分布的去中心化，Shannon 熵即时上升 18.7%、HHI 即时下降 11.8%，单条话题重复上榜次数下降 23.4%、连续在榜时长缩短 31.7%，头部话题的垄断格局被显著瓦解。与此同时，分级分类治理导致社会类内容质量显著提升（信息差指数下降 20.0%、官媒占比上升 25.2%），而娱乐类质量保持稳定，形成了差异化的质量改善格局。然而，在注意力总量守恒的零和约束下（周度总热度守恒，$p = 0.626$），社会类份额即时下降 3.16 个百分点、娱乐类即时上升 3.84 个百分点，呈现出典型的“水床效应”——

监管对高风险领域的强化治理引发了注意力向低风险领域的系统性转移。在密度层面，社会类呈现“少而精”调整（话题数量减少、单条时长增加），娱乐类呈现“多且长”调整（数量和时长同步扩张），揭示了平台在宏观再配置基础上的微观精细化调整策略。

上述发现表明，算法治理并非单一维度的线性干预，而是触发了平台内容生态的系统性重构。多样性提升与质量改善的积极效果，与注意力再分配的非预期后果并存，揭示了算法治理中多元目标之间的潜在张力。

（二）理论贡献

本研究在以下三个方面推进了算法治理与平台经济的理论发展：

1. 揭示了算法治理的“水床效应”机制

既有研究多聚焦于算法治理的直接效果（如内容质量、多样性），较少关注治理引发的间接效应和非预期后果。本研究首次在实证层面揭示了算法治理的“水床效应”：在注意力总量约束下，对某一内容领域的监管强化将引发注意力向其他领域的系统性转移。这一发现拓展了平台治理的理论边界，表明监管效果的评估须超越单一维度，关注治理的系统性影响和溢出效应。

水床效应的理论意义在于：它揭示了注意力经济中“零和博弈”的深层逻辑——用户注意力作为稀缺资源，其总量在短期内近似固定，监管改变的是注意力的配置结构而非总量。这一洞见为理解算法治理的传导机制提供了新的分析框架。

2. 构建了“监管冲击—平台响应—生态重构”的因果链条

本研究构建了算法治理影响平台内容生态的完整因果链条：监管通过分级分类治理、多样性约束、质量门槛等制度安排改变平台的约束条件和激励结构；平台作为理性主体，通过算法调整、内容筛选、配额再分配等策略在新约束下最优化自身收益；平台响应进而引发内容生态的多维重构。

这一因果链条的理论贡献在于：它揭示了算法治理的“间接治理”逻辑——监管并非直接作用于内容本身，而是通过改变平台激励来间接重塑内容生态。这一逻辑既解释了算法治理的有效性（低执行成本、高覆盖范围），也揭示了其局限性（非预期后果、目标冲突）。

3. 提出了多元治理目标间的权衡框架

本研究基于实证发现，提出算法治理存在多元目标间的潜在权衡。在注意力总量相对稳定的约束下，监管者难以同时最大化以下三个目标：（1）信息多样性：打破头部垄断，提升内容分布的均匀程度；（2）内容质量：提高高风险内容的质量门槛，减少低质内容；（3）公共议题曝光：确保社会类内容获得充分的用户触达。

本研究的实证结果显示，政策在多样性和社会类质量两个维度取得了积极效果，但社会类份额的下降可能削弱公共议题的曝光机会。这一权衡框架为理解算法治理的复杂性提供了理论工具，也为政策设计提供了分析基础。

（三）政策启示

基于上述理论发现，本研究为算法治理的政策优化提出以下建议：

1. 建立多维度治理效果评估体系

当前监管评估多聚焦于单一维度（如内容质量或多样性），本研究表明算法治理具有系统性影响，须建立涵盖多样性、质量、注意力配置、用户效用等多维度的综合评估体系。具体建议包括：（1）将“注意力配置结构”纳入监管评估指标，监测不同内容类型的份额变化；（2）建立“非预期后果”预警机制，识别治理措施可能引发的溢出效应；（3）开展定期的治理效果审计，动态调整监管策略。

2. 探索差异化的正向激励机制

当前监管主要通过提高违规成本来约束平台行为，本研究发现这一“负向激励”模式可能导致水床效应。建议探索正向激励机制作为补充：（1）对高质量社会类内容给予流量扶持或经济激励，弥补监管导致的份额下降；（2）将“公共价值贡献”纳入平台评价体系，激励平台主动提升公共议题的曝光质量；（3）建立“质量—曝光”联动机制，使高质量内容获得与其公共价值相匹配的注意力配额。

3. 增强用户对算法的知情权与选择权

本研究揭示的水床效应本质上是平台在监管约束下的最优化响应，用户在此过程中处于被动地位。建议监管政策在约束平台算法的同时，推动平台向用户提供算法偏好的调整选项：（1）允许用户在一定范围内调整不同内容类型的推荐权重；（2）提供“探索模式”与“偏好模式”的切换功能；（3）增强算法推荐的透明度，让用户了解内容配置的变化及其原因。这种“用户赋权”策略既保障了平台层面的基本多样性，又尊重了用户偏好的异质性。

4. 建立跨平台协同治理机制

本研究聚焦单一平台，但用户的跨平台迁移可能削弱单一平台监管的效果。如果用户因监管导致的内容变化而转向其他平台，则单一平台的治理效果将被稀释。建议在主要平台间建立协同治理机制：（1）统一的内容质量标准和多样性要求；（2）低质内容信息共享和联合治理；（3）同步执法以降低跨平台套利空间。

（四）研究局限与未来方向

本研究存在以下局限，为未来研究提供了方向：

数据局限。第一，用户偏好数据缺失。本研究仅观测到平台供给侧的内容配置变化，无法直接测度用户需求侧的真实偏好变化，未来研究可通过用户调查或行为实验弥补这一不足。第二，跨平台数据缺乏。本研究聚焦微博单一平台，无法验证用户是否因监管而将注意力转移至其他平台，未来研究可收集多平台数据进行比较分析。第三，观测窗口有限。本研究观测窗口为政策实施后约 10 个月，仅能捕捉短期和中期效应，长期效应有待追踪。

未来研究方向。第一，开展用户层面的田野实验，向部分用户提供算法控制权，观测用户的真实选择行为和效用变化，验证“用户赋权”策略的有效性。第二，收集多平台用户行为数据，检验单一平台监管是否导致用户跨平台迁移，评估协同治理的必要性。第三，追踪监管后用户的信息获取行为、政治参与度、主观福利等指标，评估算法治理对用户长期效用的影响。第四，开展国际比较研究，对比中国、欧盟《数字服务法》、美国等不同治理模式的效果差异，识别有效治理实践的共同特征。

（五）结语

算法推荐系统作为数字时代的信息配置机制，深刻影响着公共信息传播和社会认知形成。本研究利用中国清朗行动这一准自然实验，系统评估了算法治理的多维效应，揭示了治理在改善内容生态的同时引发注意力结构性再分配的复杂图景。

“水床效应”的发现具有重要的政策含义：算法治理是一项需要在多元目标间寻求平衡的系统工程，单一维度的优化可能引发非预期后果。在约束平台算法的同时，建立多维评估体系、探索正向激励机制、增强用户知情权与选择权、推进跨平台协同治理，可能是实现更优治理效果的可行路径。

本研究为理解算法治理的复杂效应提供了新的实证证据和理论框架，也为完善平台治理体系提供了政策参考。随着算法技术的持续演进和监管实践的不断深化，算法治理研究将继续面临新的挑战和机遇。

参考文献

Berman, R. and Katona, Z. (2020). Curation algorithms and filter bubbles in social networks. *Marketing Science*, 39(2):296–316.

Boudreau, K. J. and Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. In Gawer, A., editor, *Platforms, Markets and Innovation*, pages 163–191. Edward Elgar Publishing.

Chakraborty, A., Paranjape, B., Kakarla, S., and Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 9–16. IEEE.

Gelper, S., Peres, R., and Eliashberg, J. (2018). Talk bursts: The role of spikes in prerelease word-of-mouth dynamics. *Journal of Marketing Research*, 55(6):801–817.

Golman, R. and Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*, 5(3):143–164.

Gorwa, R., Binns, R., and Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1):1–15.

Gritsenko, D. and Wood, M. (2022). Algorithmic governance: A modes of governance approach. *Regulation & Governance*, 16(1):45–62.

Hagiu, A. and Wright, J. (2015). Multi-sided platforms. *International Journal of Industrial Organization*, 43:162–174.

He, S., Hollenbeck, B., and Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Hosanagar, K., Fleder, D., Lee, D., and Buja, A. (2014). Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4):805–823.

Huber, T. L., Fischer, T. A., Dibbern, J., and Hirschheim, R. (2013). A process model of complementarity and substitution of contractual and relational governance in IS outsourcing. *Journal of Management Information Systems*, 30(3):81–114.

Jacobides, M. G., Cennamo, C., and Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8):2255–2276.

Jhaver, S., Bruckman, A., and Gilbert, E. (2019). Does transparency in moderation really matter? user behavior after content removal explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–27.

Kleinberg, J. and Raghavan, M. (2021). Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118.

Leskovec, J., Backstrom, L., and Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. pages 497–506.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1):75–98.

Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., and Lehmann, S. (2019). Accelerating dynamics of collective attention. *Nature Communications*, 10:1759.

Oestreicher-Singer, G. and Sundararajan, A. (2012). The visible hand? demand effects of recommendation networks in electronic markets. *Management Science*, 58(11):1963–1981.

Parker, G. G. and Van Alstyne, M. W. (2018). Innovation, openness, and platform control. *Management Science*, 64(7):3015–3032.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146:102551.

Song, P., Xue, L., Rai, A., and Zhang, C. (2018). The ecosystem of software platform: A study of asymmetric cross-side network effects and platform governance. *MIS Quarterly*, 42(1):121–142.

Tiwana, A., Konsynski, B., and Bush, A. A. (2010). Research commentary—platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4):675–687.

Ulbricht, L. and Yeung, K. (2022). Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation & Governance*, 16(1):3–22.

Zhu, F. and Liu, Q. (2018). Competing with complementors: An empirical look at Amazon.com. *Strategic Management Journal*, 39(10):2618–2642.